**BIOM262: ChIP-Seq workshop – Tuesday Feb 15th 2018**

We will follow the bedtools tutorial and use HOMER for studying NDRs (nucleosome depleted regions).

**BEDtools:**

We will follow the tutorial by Aaron Quinlan

<http://quinlanlab.org/tutorials/bedtools/bedtools.html>

some notes:

By default, bedtools genomecov will compute a histogram of coverage for the genome file provided. The default output format is as follows:

1. chromosome (or entire genome)
2. depth of coverage from features in input file
3. number of bases on chromosome (or genome) with depth equal to column 2.
4. size of chromosome (or entire genome) in base pairs
5. fraction of bases on chromosome (or entire genome) with depth equal to column 2.

We will try using genomecov on our data:

**bedtools genomecov -ibam /home/ucsd-train36/biom262-2018/class-iii/data/interphase-h3k27ac-chr12.bam -bg | awk '$4 > 9' > bed\_cov.bg**

**Plotting the NDR signals**

The data from Javasky & Shamir *et al*. is derived from human cells, so we first should add to HOMER the needed genome:

**perl** <path; mine was at ~/software/homer/>/configureHomer.pl **-install hg19**

The bam files for chromosome 12 are located at

/home/ucsd-train36/biom262-2018/class-iii/data/

Since Homer will want to access the files directly, make sure you copy the data directory to your account in a meaningful directory:

*cp -R /home/ucsd-train36/biom262-2018/class-iii/data <location in your directory>*

(Note, if the above does not work due to limited space, look for large data sets in your account you no longer need and can remove. Usually It’s *.sam* files)

Now, let’s make a tag-directories for all files (as we did before):

***for f in*** *<path to your bam files>****/\*.bam; do fname=`basename $f -chr12.bam`; makeTagDirectory <****path to where your tag directories will be****>/$fname -genome hg19***

***-checkGC $f; done***

Once we have the tag-directories, we can use the annotatePeaks.pl command

***annotatePeaks.pl tss hg19 -size 4000 -hist 10 -d*** *<*path to the tagdirs>/\*k27\* > <path>/h3k27ac-hist.txt

and

***annotatePeaks.pl tss hg19 -size 4000 -hist 10 -d*** *<*path to the tagdirs>/\*k4\* > <path>/h3k4me3-hist.txt

Open the files using a spreadsheet program (R, Excel or similar). As we discussed, the first column gives the distance offsets from the TSS followed by columns corresponding to the ‘coverage’ (calculated by extending tags by their estimated ChIP-fragment length; analogous to the profiles made for the Genome Browser), ‘+ Tags’, and ‘- Tags’ (density of 5’ and 3’ aligned tags; these are independent of fragment length) for each experiment.

Plot the coverage of mitosis and interphase H3K27ac and H3K4me3 – do you see a difference in the way they are organized around the TSS?